BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene

Neeme Tõnisson*†, Jana Zernant*, Ants Kurg†, Hendrik Pavel*, Georg Slavin*, Hanno Roomere*†, Aune Meiel*†, Pierre Hainaut‡, and Andres Metspalu*†‡§

*Asper, Ltd., 3 Oru Street, 51014 Tartu, Estonia; †Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, 23 Riia Street, 51010 Tartu, Estonia; and ‡International Agency for Research on Cancer, 150, Cours Albert Thomas, F-69372 Lyon Cedex 08, France

Communicated by C. Thomas Caskey, Cogene Biotech Ventures, Ltd., Houston, TX, February 20, 2002 (received for review July 9, 2001)

Identification of mutations in the tumor suppressor gene TP53 has implications for the molecular epidemiology and for the molecular pathology of human cancer. We have developed and evaluated an arrayed primer extension assay for covering both strands of a region of the coding sequence containing more than 95% of the mutations described so far in TP53. On average, 97.5% of the arrayed TP53 gene sequence can be analyzed from either sense or antisense strands, and 81% from both strands. A patient DNA sample is amplified and annealed to arrayed primers, which then promote DNA polymerase extension reactions with four fluorescently labeled dideoxynucleotides. The TP53 gene chip spans exons 2–9 plus two introns from both strands. The performance of the assay was evaluated by using freshly extracted genomic DNA, as well as DNA extracted from archival (paraffin-embedded) DNA samples. The arrayed primer extension-based TP53 gene test provides an accurate and efficient tool for DNA sequence analysis of this frequently mutated gene for both research and clinical applications.

APEX | oligonucleotide array | chip

The evidence is growing that specific mutations in the TP53 gene can represent important factors for the prognosis of cancer and for the response to various types of cytotoxic therapy. Furthermore, patterns of TP53 mutations have differed considerably from one type of cancer to the other (1–4). However, screening for TP53 mutations gene has yet to become a routine in clinical or epidemiological practice, mainly because current detection technologies are labor-intensive and have prohibitive costs for large-scale prospective studies. Another strong limitation to routine analysis of TP53 mutations resides in the fact that many tumors contain an excess of wild-type TP53 as compared with mutant, resulting from the presence of intact alleles in tumor as well as in noncancer cells (stroma, inflammatory cells, blood vessels).

In this report we describe the development of an arrayed primer extension (APEX) assay for the rapid and sensitive detection and identification of mutations in the TP53 gene. APEX is a genotyping and resequencing technology that combines the advantages of Sanger dideoxy sequencing with the parallelization and high-throughput potential of the microarray format. A DNA sample is amplified, fragmented enzymatically, and annealed to arrayed primers, promoting sites for template-dependent DNA polymerase extension reactions by using four fluorescently labeled dideoxynucleotides. Each base is probed with two primers, one for the sense and another for the antisense strand (5). GENORAMA imaging system and genotyping software (Asper Ltd., Tartu, Estonia, www.asperbio.com) were used for imaging and semiautomatic sequence analysis (Fig. 1).

The principle of sequencing by primer extension on oligonucleotide array has been successfully applied for the systematic identification of all common TP53 mutations in human cancers. The TP53 microarray presented here spans exons 2–9 [containing more than 98% of all mutations described so far in human cancer (6)], together with flanking splice sites and introns 5 and 8 from both strands (total of 1,218 bases; Fig. 2). This system has been designed to allow the detection of most common mutations (missense, nonsense, tandem, insertions, deletions, and complex mutations) and all identified polymorphisms in the TP53 coding sequence. We found that this system allows for sequencing of an average of 97.5% of the arrayed TP53 gene from either sense or antisense strand, whereas 81% of the whole sequence was simultaneously analyzed from both strands. The length of this simultaneous DNA sequence readout (1.2 kb from both strands) outmatches the limits of the current standard for mutation detection, automated dideoxy sequencing. We describe performance of this assay, evaluated by using 100 normal genomic DNA samples from the Estonian population, plus DNA extracted from 11 archival pathology sections (paraffin-embedded resections of primary esophageal cancers), which were demonstrated to contain TP53 mutations by using classical mutation detection methods [temporal temperature gradient gel electrophoresis (TTGE), followed by direct sequencing].

Two silent, six missense, one splice-site mutation, and an insertion were confirmed by both techniques (Table 1). One of the tumors showed a missense mutation at codon 290 by APEX, instead of a silent, point mutation as detected by TTGE plus dideoxy sequencing. In addition, one point mutation, which escaped detection by TTGE plus dideoxy sequencing, was identified by APEX. On the basis of these results, we conclude that the APEX-based TP53 mutation assay provides an accurate and cost-efficient tool for DNA sequence analysis of this frequently mutated gene. Additional oligonucleotides or regions of the TP53 gene can be easily added to the assay. This prototypic assay represents a valuable platform for the development of diagnostic sequencing assays, for TP53 and other genes of interest.

Methods

Template Preparation. Exons 2–9 of the TP53 gene were amplified from genomic DNA in three amplicons: exons 2–4 (with 5′-TGGAAGTGTCTCATGCTGGA and 5′-ATACGGCCAGGCATTGAAGT primers), exons 5–6 (with 5′-TCTGTCTCCTTCCTCTTCCT and 5′-CACTGACAACCACCCTTAAC primers), and exons 7–9 (with 5′-CTCATCTTGGGCCTGTGTTA and 5′-GCCCCAATTGCAGGTAAAAC primers). A 20% fraction of the dTTP in the amplification mixture was substituted by dUTP (5, 7). The amplification products were concentrated and purified by ethanol precipitation in the presence of ammonium acetate. Fragmentation and functional inactivation of the unincorporated dNTPs was achieved in a one-step reaction by addition of shrimp alkaline phosphatase (Amersham Biosciences, Piscataway, NJ) and thermolabile ura-

GENETICS

Abbreviations: APEX, arrayed primer extension; TTGE, temporal temperature gradient gel electrophoresis.

§To whom reprint requests should be addressed. E-mail: andres@ebc.ee.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

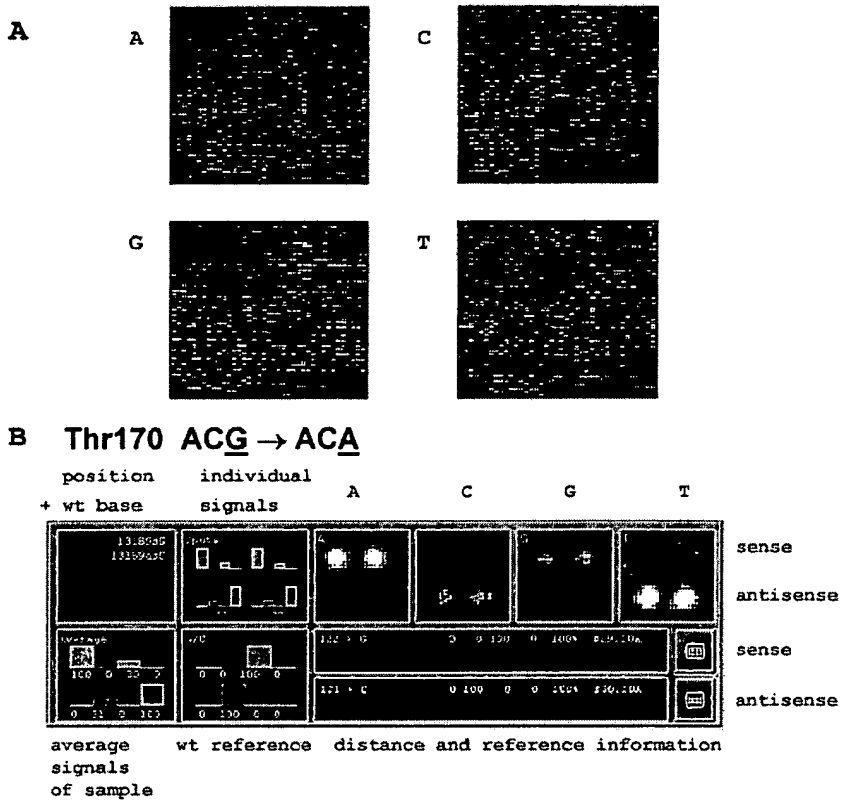

Fig. 1. TP53 APEX-based sequencing assay. (*A*) Grayscale images for each fluorescent dideoxy nucleotide are used for the sequence analysis. (*B*) Silent mutation in the third base of codon 170 of TP53, analyzed by the GENORAMA software. Signals from the analyzed base are averaged and the signal pattern obtained is compared with the wild-type (wt) reference. Grayscale bitmaps corresponding to all four fluorescent dideoxy nucleotides at the base to be determined are shown enabling visual analysis. A signal in the sense area and T in the antisense area are indicative for mutation in the current tumor sample. The distance and reference information consist of: (*i*) the distance measure of the given signal pattern from the wt reference pattern; (*ii*) the wt base with relative signal intensities at four (A, C, G, and T) fluorescence channels; (*iii*) percentage of the signal pattern at the wt reference cluster database for the given base; and (*iv*) index given by the GENORAMA software.

cil *N*-glycosylase (Epicentre Technologies, Madison, WI) (5) and heat treatment.

Oligonucleotide Microchips. APEX primers were designed, according to the wild-type sequence of the human TP53 gene (accession no. U94788) for both sense and antisense directions. The 25-mer oligonucleotides with 12-carbon amino linkers at their 5′ end were obtained from Genset (Paris). Used for spotting the oligonucleotides were 24 × 60 mm aminosilane plus phenylene diisothiocyanate-coated microarray slides (8) (Asper, Ltd.).

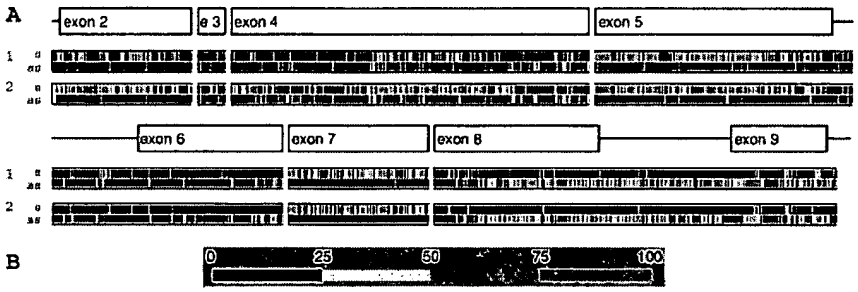

Fig. 2. Performance of the APEX-based sequencing in different regions in the TP53 gene. In some regions of the gene, sense and antisense strands have different performance in APEX. For instance, in exon 7, which can be viewed as an extreme case, the antisense strand signals are very good. At the same time the signals from the sense strand perform below the average level. (*A*) Performance of oligos corresponding to different regions of the p53 gene from sense (s) and antisense (as) strands is shown with data from two independent series of experiments. The upper bar (1) represents APEX performance from 20 repeated APEX reactions with the same wild-type reference DNA, whereas only the automatically clustered signal intensities are used. The lower bar (2) represents visually corrected data from 100 healthy individuals sequenced by APEX. Both of the patterns are highly overlapping. (*B*) Color code and scale for the image. Oligos with signals matching the wt sequence at least in 75% of experiments are shown black. Oligos with zero signals or signals different from the wt sequence are shown yellow, orange, or red.

Table 1. Mutations detected by TTGE and APEX assays

Sample	Mutation: Codon	Mutation: Nucleotide	Mutation: Amino acid	TTGE + sequencing	APEX
8	172	GTT→TTT	Val→Phe	+	–
	315	TCT→TGT	Ser→Cys	–	+
10	175	CGC→CAC	Arg→His	+	+
13	213	CGA→CGG	Arg	+	+
15	170	ACG→ACA	Thr	+	+
16	Intron 5 splice	TAG→TAA		+	+
20	179	CAT→CGT	His→Arg	+	+
22	170	ACG→ACACG	ins 2 bp	+	+*
25	273	CGT→TGT	Arg→Cys	+	+
31	290	CGC→CGA	Arg	+	–
	290	CGC→CCC	Arg→Pro	–	+
48	164	AAG→ACG	Lys→Thr	+	–
	286	GAA→AAA	Glu→Lys	+	+
53	175	CGC→CAC	Arg→His	+	+
	213	CGA→CAA	Arg→Gln	+	+
11		Total		13	12

Concordance with TTGE plus dideoxy sequencing as the reference was 10 of 13. One mutation was identified in the same codon by APEX compared with dideoxy sequencing (sample 31). One mutation was identified by APEX only (sample 8).

*Presence of mutated base (A) determined from the sense strand only.

Primers were diluted to 50 μM concentration in 100 mM carbonate buffer, pH 9.0, and spotted onto the activated surface with Affymetrix 417 arrayer (Affymetrix, Santa Clara, CA). The slides were blocked with 1% ammonia solution and stored at 4°C until needed. Washing steps with 95°C water and 100 mM NaOH were performed before APEX reactions to reduce the background fluorescence and avoid rehybridization of unbound oligonucleotides to the APEX slide.

Genomic DNA Samples from Estonian Population. The genomic DNA samples from healthy individuals were obtained from the Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, and comprised a subset of samples collected within the framework of the project The Influence of Genetical and Environmental Factors on Health of Estonian Population of the Estonian Ministry of Social Affairs. The project had been approved by the ethics committee of University of Tartu. Informed consent was signed by all the participants of the study.

APEX-Based Resequencing. One-third of a product from 50 μl of PCR was used for each primer extension reaction. The APEX mixture consisted of 10 μl of fragmented product, 4 units of Thermo Sequenase DNA polymerase (Amersham Pharmacia), 2 μl of Thermo Sequenase reaction buffer (260 mM Tris·HCl, pH 9.5/65 mM $MgCl_2$) (Amersham Pharmacia), and 2 μM final concentration of each fluorescently labeled ddNTP: Texas Red-ddATP, Cy3-ddCTP, fluorescein-ddGTP, Cy5-ddUTP (Amersham Pharmacia; NEN). The DNA in buffer was denatured at 95°C for 5 min. The enzyme and dye terminators were immediately added to other components, and the whole mixture was applied to prewarmed slides at 58°C. The reactions were allowed to proceed 20 min under parafilm and stopped by washing at 95°C for 2 × 90 s in MilliQ water. A droplet of SlowFade Light Antifade Reagent (Molecular Probes) was applied to the microchips to limit bleaching of the fluorescein. The slides were imaged with the Genorama imaging system (Asper, Ltd.), at 20-μm resolution.

The TP53 gene sequence and mutations were identified by GENORAMA 3.0 genotyping software by using clustered signal patterns from a sequenced wild-type DNA as the statistical reference. The distances of signals from the clusters were used as measures of match with the wild-type gene sequence. Distance (d) of the sample signal pattern compared with the signal patterns in the wild-type reference cluster database were calculated as follows:

$$d = \sqrt{\sum (N_c - N_s)^2},$$

where N_c is the signal intensity of the given nucleotide (A, C, G, and T) in the cluster database, and N_s is the signal intensity of the given nucleotide of the DNA sample.

Results

Oligonucleotide Design. Each base in TP53 is identified by two unique 25-mer oligonucleotides, one for sense and one for antisense strand (total of 2,436 oligonucleotides for the analyzed sequence). The oligonucleotides are based on TP53 wild-type sequence (accession no. U94788), with their 3′ends one base upstream of the base to be identified. The vast majority of these oligonucleotides performs well in APEX. A fraction of the oligonucleotides formed secondary structures, either enabling signals from self-priming or interfering with annealing to test DNA, and therefore needed redesigning. Although the 3′ end and its proximity of the primers cannot be modified, the internal part of the primer may be changed by incorporating a mismatch without seriously affecting the target-specific priming ability. Oligonucleotides for 5.9% of the sequenced bases from either strand were redesigned by introducing a mismatch to reduce the stability of the predicted dimers and avoid self-priming. After modification, 62% of these oligonucleotides generated signals only in the presence of target DNA and not from oligonucleotide dimers; 21% of the modified oligonucleotides did not give any signal either from the target DNA or self-priming because of their reduced hybridization ability; 17% of the modified sequences produced weak or undetectable signals in half of the experiments. None of the modified oligonucleotides generated false-positive signals in the absence of the target DNA.

Some areas of the gene are difficult to sequence from both strands (Fig. 2*A*) for multiple reasons, including sequence repeats, regions with very high GC content, sequences corresponding to oligonucleotide with AT-rich 3′ ends, etc. However, only a very limited number of bases (2.5% on average) were not detected from either strand at the present state of assay development.

Sequence Analysis Algorithm. As a general strategy in APEX, the sequence can be identified either from a single experiment or interpreted on the basis of a statistical analysis. Statistical analysis facilitates identification of deviations from the wild-type reference signal pattern indicative of mutations (Fig. 1*B*). The level of possible secondary signals in the wild-type reference is useful for determining a threshold for acceptance or rejection of signals interpreted as mutations. A sequenced wild-type genomic DNA from a healthy individual was used to create a reference database of signal patterns. The signals from all of the oligonucleotides were analyzed by a clustering algorithm, grouping the signal patterns from four fluorescence channels. Each base in the sample was compared with the wild-type reference, and the value of the distance (see *Methods*) between the signal pattern and the corresponding wild-type base was used as a measure for calling the given base. Zero distance indicates a perfect match between the given base and the wild-type reference base. The analysis was performed in an automated manner, and only a subset of signals needed visual examination.

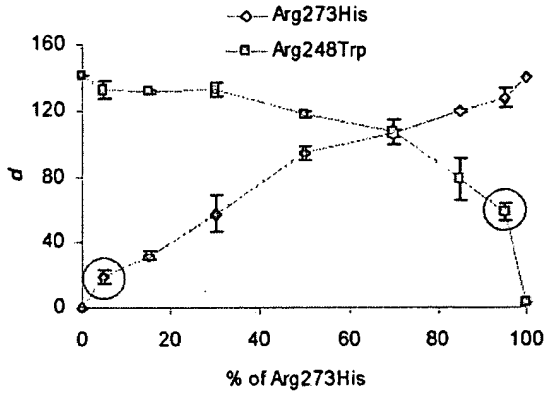

Fig. 3. APEX sensitivity for the fraction of mutated DNA. Relationship between the fraction of mutated DNA and the value of distance measure from the wild-type reference signal pattern. PCR products from two cDNA clones with known missense mutations Arg273His (CGT→CAT) and Arg248Trp (CGG→TGG) were used for the titration. Both mutations are analyzed at the DNA strand with G to A change. Five percent content of mutated DNA is detectable in both cases (indicated by circles). Error bars represent the standard deviations.

APEX Performance in TP53 Sequencing Tests with Numerous Samples. To evaluate performance of the TP53 APEX assay in large-scale studies, 100 normal DNA samples from the Estonian population were tested for common, single-nucleotide polymorphisms and for possible point mutations. A common single-nucleotide polymorphism in exon 4 (Arg-72 → Pro; Arg72Pro) was found with minor allele frequency of 0.26. The identified single-nucleotide polymorphism matches the Hardy–Weinberg equilibrium by the calculated χ^2 value ($P > 0.05$). We also detected two silent point mutations in codons 36 (CCG to CCA) and 139 (AAG to AAA) (6) in two analyzed samples. The first one may correspond to a rare polymorphism, which has been identified in up to 4% of the general population (9). At present, no evidence has been published regarding the status of silent mutation at codon 139, but it cannot be ruled out that it might also correspond to a previously unrecognized, rare polymorphism.

On average, 97.5% of the arrayed TP53 sequence was identified in our current version of the TP53 assay from either sense or antisense strand, and 81% from both strands. In the best cases, respectively, up to 99.8% and 96% of the sequence were analyzed.

Sensitivity for Mutated DNA. DNA extracted from tumor samples always contains a background of normal DNA. APEX sensitivity for the minimal identified percentage of mutated DNA was titrated by mixing PCR products obtained from the mutant (Arg248Trp and Arg273His) TP53 cDNA clones at different ratios (Fig. 3). The mutations are located in different exons, and the clones were therefore used as a competitor fraction of normal DNA for each other. The signal patterns were different from the wild type, and both mutations were detected even if the sample contained as little as 5% of mutant DNA. The samples with zero percent of mutated DNA were matching the wild-type reference DNA (Arg273His with zero distance and Arg248Trp with a distance value of 3). The mixed samples with 5% of mutated DNA, on the contrary, did not match the signal pattern of the reference wild-type sample (average distances, 19 for Arg273His and 58 for Arg248Trp). In fact, 5% of the mutated DNA allowed identification from the analysis software window (Fig. 1*B*) by eye. APEX sensitivity to detect deletions was titrated with del 13–19 TP53 cDNA clone. The first base after deletion is detected instead of the first deleted base (10, 11). The deletion was detected with sensitivity equal to a point mutation

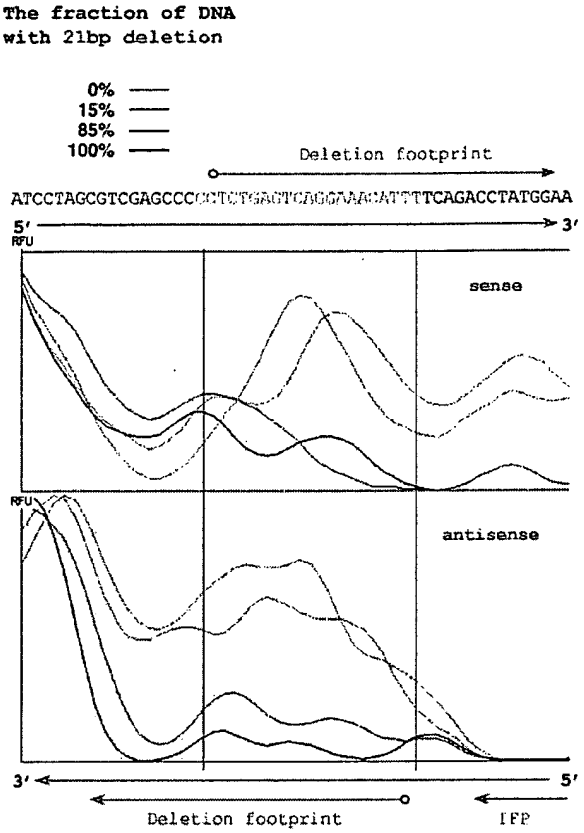

Fig. 4. APEX sensitivity for detecting deletions. Detailed patterns of signal intensities in the deletion area. The actual footprint (with weaker or missing signals) exceeds the deleted sequence by 13 to 15 bases in the 3′ direction from either strand because of partial annealing with the target sequence. The first base after deletion is detected instead of the first deleted base. Because of cDNA used as the template, intron 3 footprint (IFP) is detectable in the antisense strand.

by analyzing first base after the deletion. The complementing algorithm, based on detection of decreased signal intensities and deletion-specific footprint was less sensitive and required at least 15% mutant sequence for detection (Fig. 4). The actual footprint (with missing or weaker signals) exceeds the deletion by 13–15 bases at 3′ direction from either strand because of the partial annealing with the target sequence (Fig. 4).

Blind Test with Tumor Samples. The tumors tested were from a series of squamous cell carcinomas of the esophagus collected in Iran between 1992 and 1998. These cancers often contain TP53 mutations and are very good examples of a type of cancer in which TP53 mutation analysis may have a strong impact in clinical and epidemiological applications. Eleven samples, with a total of 12 point mutations and a 2-bp insertion in TP53, previously identified by TTGE plus manual or automated dideoxy sequencing of the extracted heteroduplex band, were used in a blind test for sensitivity and accuracy of APEX. Sequencing of the heteroduplex band has superior sensitivity to direct sequencing but requires gel purification of the PCR product. The total number of mutations determined was similar in both techniques. Two silent, six missense, one splice site mutation, and an insertion were concordantly identified (Table 1). A missense mutation at the codon 290 was found by APEX instead of a silent point mutation as identified by TTGE plus

dideoxy sequencing. One missense mutation not previously identified by TTGE plus sequencing was *de novo* identified by APEX. Only wild-type APEX signals were present in two samples, where missense mutations were previously determined.

Discussion

A practical approach to TP53 mutation screening has to combine affordable cost, high throughput, high specificity, and high sensitivity. So far, the most advanced, current alternative to dideoxy sequencing is the GENECHIP p53 assay (Affymetrix, www.affymetrix.com), which has been recently evaluated (12–14). The Affymetrix chip has good overall performance but a limited ability to detect deletions and insertions. Promising efforts have been made to couple the oligonucleotide array technology to single-base extension reaction by the DNA polymerase (10). Another recent approach, pyrosequencing, has shown accurate results for detection of mutations in a few exons of TP53 (15).

The currently described APEX-based sequencing approach by comparing a sample with the wild-type reference by the distance measure is comparable with the GENECHIP p53 assay where a score from a mixture of variables between the wild-type reference and a given sample is calculated. The higher the score for a probe set contributing to a given base, the higher the likelihood for the base being mutated (12, 14). In the GENECHIP p53 assay, the single cutoff level for calling mutations has been reported to be unsatisfactory (14). The same situation could apply to the TP53 APEX-based sequencing assay, but further studies are needed to evaluate the possible benefit of approaching each base as a separate entity. The applicable cutoff value for base calling also depends on whether the sample is analyzed for germ-line or somatic mutations. In the current work, prescan of the sequence was made with a general cutoff distance. The positions exceeding the threshold distance from the wild-type signal pattern were visually verified. Just one APEX oligonucleotide per each sequenced base and the general low noise makes possible the fast visual inspection at positions where the software is giving ambiguous results.

The results from the 100 healthy individuals analyzed are encouraging for applying APEX in large-scale TP53 studies, whereas single-nucleotide polymorphism data can have an impact on the analysis of individual risks or of cancer outcome. The identified Arg72Pro polymorphism has recently been proposed to play a role in tumorigenesis. Controversial evidence exists that the Arg-72 allele might be more sensitive to degradation induced by the oncoproteins of human papilloma viruses, suggesting that this polymorphism may predispose to cervical cancer (16). On the other hand, recent studies have shown that the cellular interactions of mutant p53 protein may be different depending of the allelotype of codon 72 (17). The fact that our assay can simultaneously perform mutation detection and correct identification of codon 72 status adds further weight to its usefulness as a one-step assay in clinical or epidemiological studies.

The TP53 detection limit for known alleles was identified as low as in 5%. The actual limit could sometimes be even less than 5%, but in real life the possible alleles are mostly unknown and reliable control and comparison with results obtained with standard methods can be technically difficult because of their own error rates. The dideoxy chain termination sequencing (ref. 18; Fig. 5) and the pyrosequencing (15) are operating at a 30% detection limit of mutation-specific signals. Heteroduplex analysis techniques like TTGE have 10^{-2} sensitivity under optimized conditions (19), but the most commonly used screening method, single-strand conformational polymorphism, has been shown to produce also 5% false-positive results (20). A potential explanation is misincorporation of bases in PCR. Therefore, the fraction of mutated DNA was not further diluted, and the TP53 APEX-based sequencing was evaluated with tumor samples in a blind test.

The total number of mutations determined in 11 esophageal cancer samples was similar by both APEX and TTGE plus

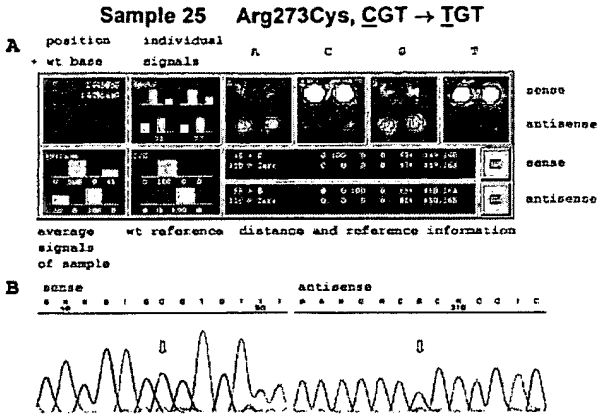

Fig. 5. Missense mutation Arg273Cys CGT→TGT, difficult to identify by automated dideoxy sequencing. (*A*) First base of TP53 codon 273, analyzed by APEX-based sequencing. The signals corresponding to T in the sense strand and A in the antisense strand are indicative for mutation. (*B*) Automated dideoxy sequencing images corresponding to the Arg273Cys mutation from both DNA strands. The indicated mutation-specific peaks are in the range of the background noise and can be easily missed by visual analysis.

sequencing. Two silent, six missense, one splice-site mutation, and an insertion were concordantly identified (Table 1). A missense mutation at the codon 290 was found by APEX instead of a silent-point mutation as identified by TTGE plus dideoxy sequencing. One missense mutation not previously identified by TTGE plus sequencing was *de novo* identified by APEX. Two samples with missense mutations escaped identification by APEX. However, in these specimens, identification of the mutation was possible only by dideoxy sequencing of a PCR product generated from excised TTGE bands with abnormal migration patterns, indicating that mutant DNA was present only in a tiny fraction of the tumor. The latter results suggest that performance and sensitivity of the APEX-based sequencing could be enhanced and all of the mutations possibly identified by use of enrichment techniques such as microdissection of tumor cells from the sample.

In conclusion, we have developed and evaluated an APEX-based sequencing test at the scale of the almost complete TP53-coding sequence, providing an accurate and cost-efficient tool for DNA sequence analysis of this frequently mutated gene. Novel analysis algorithms were developed enabling automatic sequencing. The evaluation test with tumor samples showed performance comparable with one of the most sensitive and also laborious technologies available, dideoxy sequencing of heteroduplex band obtained by TTGE. However, due to the reduced number of steps in template preparation and the possibility of performing automated analysis, APEX is much more suitable for developing tests for high-throughput in clinical diagnostics and large scale epidemiological studies.

Note Added in Proof. When this manuscript was in process, a paper describing the resequencing of exon 7 in the TP53 gene was published (21).

We thank Dr. A. Kristjuhan and V. Jaks, who kindly provided the mutant TP53 cDNA clones; E. Haamer, L. Land, I. Valvas, and V. Soo for technical assistance; Dr. K. Kask for critical reading of the manuscript; and Mrs. Krista Liiv for superior organizing work throughout the whole project. This work was supported in part by research grants from the European Community (IC15-CT98-0309), the Estonian Science Foundation (4479), and the Estonian Ministry of Education (Core Grant 0181518s98). A. Metspalu received the Visiting Scientist Award of the International Agency for Research on Cancer, Lyon, France.

1. Thorlacius, S., Borresen, A. L. & Eyfjord, J. E. (1993) *Cancer Res.* **53**, 1637–1641.
2. Wen, W. H., Reles, A., Runnebaum, I. B., Sullivan-Halley, J., Bernstein, L., Jones, L. A., Felix, J. C., Kreienberg, R., el-Naggar, A. & Press, M. F. (1999) *Int. J. Gynecol. Pathol.* **18**, 29–41.
3. Aas, T., Borresen, A. L., Geisler, S., Smith-Sorensen, B., Johnsen, H., Varhaug, J. E., Akslen, L. A. & Lonning, P. E. (1996) *Nat. Med.* **2**, 811–814.
4. Cabelguenne, A., Blons, H., de Waziers, I., Carnot, F., Houllier, A. M., Soussi, T., Brasnu, D., Beaune, P., Laccourreye, O. & Laurent-Puig, P. (2000) *J. Clin. Oncol.* **18**, 1465–1473.
5. Kurg, A., Tõnisson, N., Georgiou, I., Shumaker, J., Tollett, J. & Metspalu, A. (2000) *Genet. Test* **4**, 1–7.
6. Hernandez-Boussard, T., Rodriguez-Tome, P., Montesano, R. & Hainaut, P. (1999) *Hum. Mutat.* **14**, 1–8.
7. Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M. & Miyada, C. G. (1996) *Hum. Mutat.* **7**, 244–255.
8. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994) *Nucleic Acids Res.* **22**, 5456–5465.
9. Felix, C. A., Brown, D. L., Mitsudomi, T., Ikagaki, N., Wong, A., Wasserman, R., Womer, R. B. & Biegel, J. A. (1994) *Oncogene* **9**, 327–328.
10. Head, S. R., Rogers, Y. H., Parikh, K., Lan, G., Anderson, S., Goelet, P. & Boyce-Jacino, M. T. (1997) *Nucleic Acids Res.* **25**, 5065–5071.
11. Tõnisson, N., Kurg, A., Lõhmussaar, E. & Metspalu, A. (2000) in *Microarray Biochip Technology*, ed. Schena, M. (Eaton Publishing, Natick, MA), pp. 247–263.
12. Ahrendt, S. A., Halachmi, S., Chow, J. T., Wu, L., Halachmi, N., Yang, S. C., Wehage, S., Jen, J. & Sidransky, D. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 7382–7387.
13. Wen, W. H., Bernstein, L., Lescallett, J., Beazer-Barclay, Y., Sullivan-Halley, J., White, M. & Press, M. F. (2000) *Cancer Res.* **60**, 2716–2722.
14. Wikman, F. P., Lu, M. L., Thykjaer, T., Olesen, S. H., Andersen, L. D., Cordon-Cardo, C. & Orntoft, T. F. (2000) *Clin. Chem.* **46**, 1555–1561.
15. Garcia, C. A., Ahmadian, A., Gharizadeh, B., Lundeberg, J., Ronaghi, M. & Nyren, P. (2000) *Gene* **253**, 249–257.
16. Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I. M., Matlashewski, G. & Banks, L. (1998) *Nature (London)* **393**, 229–234.
17. Tada, M., Furuuchi, K., Kaneda, M., Matsumoto, J., Takahashi, M., Hirai, A., Mitsumoto, Y., Iggo, R. D. & Moriuchi, T. (2001) *Carcinogenesis* **22**, 515–517.
18. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R. & Chen, S. M. (1997) *Nucleic Acids Res.* **25**, 4500–4504.
19. Bjorheim, J., Lystad, S., Lindblom, A., Kressner, U., Westring, S., Wahlberg, S., Lindmark, G., Gaudernack, G., Ekstrom, P., Roe, J., *et al.* (1998) *Mutat. Res.* **403**, 103–112.
20. Yuan, B., Thomas, J. P., von Kodolitsch, Y. & Pyeritz, R. E. (1999) *Hum. Mutat.* **14**, 440–446.
21. Shumaker, J. M., Tollet, J. J., Filbin, K. J., Montague-Smith, M. P. & Pirrung, M. C. (2001) *Bioorg. Med. Chem.* **9**, 2269–2278.

GENETIC TESTING
Volume 4, Number 1, 2000
Mary Ann Liebert, Inc.

Arrayed Primer Extension: Solid-Phase Four-Color DNA Resequencing and Mutation Detection Technology

ANTS KURG,[1] NEEME TÕNISSON,[1,4] IOANNIS GEORGIOU,[2] JOHN SHUMAKER,[3] JEFF TOLLETT,[3] and ANDRES METSPALU[1,4]

ABSTRACT

The technology and application of arrayed primer extension (APEX) is presented. We describe an integrated system with DNA chip and template preparation, multiplex primer extension on the array, fluorescence imaging, and data analysis. The method is based upon an array of oligonucleotides, immobilized via the 5′ end on a glass surface. A patient DNA is amplified by PCR, digested enzymatically, and annealed to the immobilized primers, which promote sites for template-dependent DNA polymerase extension reactions using four unique fluorescently labeled dideoxy nucleotides. A mutation is detected by a change in the color code of the primer sites. The technology was applied to the analysis of 10 common β-thalassemia mutations. Nine patient DNA samples, each of which carries a different mutation, and four wild-type DNA samples were correctly identified. The signal-to-noise ratio of this technology is, on the average, 40:1, which enables the identification of heterozygous mutations with a high confidence level. The APEX method can be applied to any DNA target for efficient analysis of mutations and polymorphisms.

INTRODUCTION

GENETICS AND MOLECULAR MEDICINE have an expanding need for rapid genotyping, mutation analysis, and DNA resequencing technologies that have a clear potential for miniaturization, parallelization, and automation and enable high throughput and ability to identify changes precisely in the patient DNA. Conventional methods for mutation detection, such as single strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), chemical cleavage, or direct sequencing are time and labor intensive (Cotton *et al.*, 1998) with little parallelization. A promising solution for this technological need is the use of oligonucleotide arrays using nucleic acid hybridization (Chee *et al.*, 1996; Cronin *et al.*, 1996; Hacia *et al.*, 1996) or hybridization coupled with an enzyme-mediated reaction, either by primer extension (Shumaker *et al.*, 1996; Head *et al.*, 1997; Pastinen *et al.*, 1997) or ligation (Landegren *et al.*, 1998). The most developed approach today, hybridization of labeled target to high-density oligonucleotide microarrays (*e.g.*, Affymetrix GeneChip™ arrays), is a revolutionary method for DNA sequence analysis. Feasibility studies of this approach are promising and the first results are impressive. However, the high complexity of the assays due to the large number of oligonucleotide probes per target sequence base, sensitivity to hybridization conditions, complicated data analysis, and high cost are driving several research groups to look for alternative technologies.

Here we present Arrayed Primer EXtension (APEX) technology as an alternative to array-based DNA sequence analysis by hybridization. We describe an integrated system with chip and template preparation, multiplex primer extension on the array, fluorescence imaging, and data analysis. The method is based upon a two-dimensional (2D) array of oligonucleotides, immobilized via the 5′ terminal amino group onto an epoxysilanized glass support. This method can be viewed as dye-terminator sequencing of DNA, but instead of using one primer and analyzing hundreds of extension products in polyacrylamide gel electrophoresis (PAGE), we can use hundreds to thousands of primers that are spatially separated and extend each by only one dye-labeled nucleotide. The single-tube sample preparation protocol consists of PCR amplification followed by a DNA fragmentation reaction. After hybridization of the

[1]Institute of Molecular and Cell Biology, Tartu Childrens Hospital, University of Tartu, Estonian Biocentre, Tartu 51010, Estonia.
[2]Medical School, University of Ioannina, Ioannina 45500, Greece.
[3]Baylor College of Medicine, Houston, TX, 77030.
[4]Asper Ltd., Tartu 51014, Estonia.

target DNA to the array, target-dependent oligonucleotide extension by a DNA polymerase is used to incorporate fluorescently labeled dideoxy terminators onto the primers (Head *et al.*, 1997).

We have developed a total internal reflection fluorescence (TIRF) excitation mechanism combined with a charge coupled device (CCD) detector for high-throughput image acquisition. The signals from the spectrally separated dyes are resolved to better than 2% cross-talk with dual selection by laser excitation and bandpass filtering of the emitted fluorescence. Imaging is followed by a software analysis to convert the fluorescence information into sequence data. The APEX method combines the advantage of both high information content of the oligonucleotide array and fidelity of the enzymatic primer extension reaction. The enzyme acts as a biological proofreading mechanism, discriminating against 3′ end mismatches. Moreover, by using primer extension, each position on the array identifies a unique base of the sequence as the result of the direct competition of four dye-terminators for the same spot, thus reducing the array complexity by at least a factor of four. Signal-to-noise ratio is improved by the added fidelity of the polymerase. The elimination of intensity comparisons across multiple spots, as is the case for hybridization assays, makes the analysis more robust. Two unique oligonucleotide primers probe the sense and antisense target strands at the same base location. APEX can be used to analyze known point mutations, deletions, and insertions and can identify the presence of unknown polymorphisms.

We applied APEX for the identification of 10 common point mutations in the human β-globin gene (Cao *et al.*, 1997; Huisman *et al.*, 1998), causing β-thalassemia in their homozygous state. β-Thalassemia is a very common autosomal recessive disorder in populations of Mediterranean, Middle Eastern, and Far Eastern descent. It has been estimated that approximately 240 million people worldwide are heterozygotes for β-thalassemia and at least 200,000 affected individuals are born annually (Cao *et al.*, 1997). The β-globin gene is rather short for a human gene (~1.5 kb), but harbors more than 150 mutations worldwide. Despite this heterogeneity, each at-risk population has its own spectrum of 5–10 common mutations (Cao *et al.*, 1997). Due to this phenomenon, an array for identification of 10 common mutations would offer a platform for high-throughput genetic testing of β-thalassemia in a given population. The mutations studied in this work are typical for the Mediterranean region (see Table 1). Extra primers are needed to expand the 10-mutation-specific chip presented here to a human β-globin gene resequencing chip, allowing the analysis of nucleotide changes regardless of patient origin or mutation location.

In this report, we describe the APEX technology and present the results of a β-thalassemia mutation study. Nine DNA samples from patients and carriers, each of which carries a different mutation and four wild-type DNA samples, were correctly identified.

MATERIALS AND METHODS

DNA sequencing

DNA samples were previously sequenced by the Sanger dideoxy chain-termination method.

Template DNA preparation

A 1,420-bp fragment of the human β-globin gene (NID g183829; accession no. M36640) was amplified from human genomic DNA using the following primer sequences:
Forward primer 5′-ACAGGTACGGCTGTCATCAC-3′;
Reverse primer 5′-AGAATAATCCAGCCTTATCCC-3′.

TABLE 1. TEN COMMON POINT MUTATIONS FROM THE β-GLOBIN GENE FOR IDENTIFICATION WITH APEX

Mutation	*Sequence change*	*Class*	*Origin*	*Primers (sense/antisense)*
−87	C → G	Transcriptional mutant	Mediterranean	5′-TAGACCTCACCCTGTGGAGCCACAC 5′-CTGGGAGTAGATTGGCCAACCCTAG
Codon 5	ΔCT	Frameshift	Mediterranean Balkans	5′-ACAGACACCATGGTGCACCTGACTC 5′-CGGCAGTAACGGCAGACTTCTCCTC
Codon 6	ΔA	Frameshift	Mediterranean	5′-GACACCATGGTGCACCTGACTCCTG 5′-CAGGGCAGTAACGGCAGACTTCTCC
IVS-I-1	G → A	Splice junction change	Mediterranean,	5′-AAGTTGGTGGTGAGGCCCTGGGCAG 5′-ACCTGTCTTGTAACCTTGATACCAA
IVS-I-5	G → A	Consensus change	Mediterranean,	5′-TGGTGGTGAGGCCCTGGGCAGGTTG 5′-TTAAACCTGTCTTGTAACCTTGATA
IVS-I-6	T → C	Consensus change	Mediterranean	5′-GGTGGTGAGGCCCTGGGCAGGTTGG 5′-CTTAAACCTGTCTTGTAACCTTGAT
IVS-I-110	G → A	Internal IVS change	Mediterranean	5′-TAGGCACTGACTCTCTCTGCCTATT 5′-GCAGCCTAAGGGTGGGAAAATAGAC
Codon 39	C → T	Nonsense	Mediterranean, European	5′-GCTGCTGGTGGTCTACCCTTGGACC 5′-TCCCCAAAGGACTCAAAGAACCTCT
IVS-II-1	G → A	Splice junction change	Mediterranean, American Black, Asian	5′-GCACGTGGATCCTGAGAACTTCAGG 5′-AAACATCAAGGGTCCCATAGACTCA
IVS-II-745	C → G	Internal IVS change	Mediterranean,	5′-ATTGCTAATAGCAGCTACAATCCAG 5′-ACCATAAAATAAAAGCAGAATGGTA
Markers				5′-TTTAGCCTTAACGCCTXTGACGTCA X = A for self-extending T marker X = C for self-extending G marker, *etc.*

The PCR primers were obtained from Life Technologies, Inc. (Gaithersburg, MD). The amplification mixture was prepared and distributed into 50-μl aliquots The mixture contained: 5 μl of 10× PCR buffer (containing 200 m*M* Tris-HCl, pH 8.4, 500 m*M* KCl (Life Technologies), 2.5 m*M* $MgCl_2$, 0.25 m*M* of each deoxynucleotide triphosphate (dATP, dCTP, dGTP), 0.2 m*M* dTTP, 0.05 m*M* dUTP (Amersham Pharmacia Biotech., Inc., Milwaukee, WI), 40 pmol of each primer, and 1 unit of Platinum Taq DNA Polymerase (Life Technologies). The amplification reactions were performed in a PTC-200 instrument (MJ Research, Inc., Watertown, MA). First, an initial incubation at 94°C for 5 min. was performed, followed by 34 amplification cycles consisting of denaturation at 94°C for 30 sec; primer annealing at 61°C for 30 sec; and extension 72°C for 1 min. The final extension was at 72°C for 5 min.

The amplification products were initially concentrated and purified by ethanol precipitation in the presence of ammonium acetate. Fragmentation and functional inactivation of unincorporated dNTPs was achieved in a one-step reaction by the addition of 1/5 U of shrimp alkaline phosphatase (Amersham Pharmacia Biotech. Inc.) and 1/5 U of thermolabile uracil *N*-glycosylase (Epicentre Technologies, Madison, WI) per one amplification product. The reaction was incubated at 37°C for 1 hour and used directly in primer extension reactions.

Oligonucleotide microchips

Oligonucleotide primers were designed, according to the wild-type sequence of the human β-globin gene, for both sense and antisense directions. 25-Mer oligonucleotides with amino linkers at their 5′ ends were obtained from Genset (Paris, France). All but Codon 5 scanning oligonucleotides were designed to scan 1 bp in the wild-type sequence. To look for the Codon 5 ΔCT frameshift mutation, an antisense primer with +1 nucleotide shift in the sequence was used. Oligonucleotide primers were attached to an epoxy-activated glass surface via an amino linker at their 5′ end (Southern *et al.*, 1992; Lamture *et al.*, 1994; Shumaker *et al.*, 1996; Pastinen *et al.*, 1997). Glass slides (24 × 60-mm; Fisherfinest Premium Cover Glasses, Fisher Scientific, Pittsburgh, PA) were sonicated in acetone and 100 m*M* NaOH (5 min both), rinsed in MilliQ water, and finally sonicated for 2 min with a solution of 2% (3-glycidoxypropyl)trimethoxysilane (Gelest Inc., Tullytown, PA) in 95% ethanol solution. Unbound silane and residual water was removed by brief rinsing in 100% ethanol. Primers were diluted to 50 μM concentration in 100 m*M* NaOH and spotted onto the activated surface with the TECAN RSP 5031 pipetting robot (TECAN AG, Hombrechtikon, Switzerland) or a custom manufactured 25 Gauge, 96-tip capillary arrayer. The slides were stored in a dust-free environment at 4°C until needed and washed twice in 95°C MilliQ water prior APEX reactions. Slides prepared this way are extremely stable and can be used even after 15 months of storage.

Arrayed primer extension reactions

As estimated by comparison with a Gibco BRL mass ladder, 200–300 ng of the amplified product was used per one APEX reaction. The 20-μl primer extension reactions consisted of 10 μl of fragmented product, 4 U of Thermo Sequenase DNA polymerase (Amersham Pharmacia Biotech.), 2 μl of Thermo Sequenase reaction buffer (260 m*M* Tris-HCl, pH 9.5, 65 m*M* $MgCl_2$) (Amersham Pharmacia Biotech.), and 1 μM final concentration of each fluorescently labeled ddNTP (Amersham Pharmacia Biotech., NEN, Boston, MA). The DNA in buffer was denatured at 95°C, for 5 min. The enzyme and dye were immediately added to the other components and the whole mix was applied to prewarmed slides at 48°C. The reactions were allowed to proceed for 20 min under coverslips and stopped by washing at 80°C for 2× 90 sec in MilliQ water. A droplet of SlowFade® Light Antifade Reagent (Molecular Probes, Eugene, OR) was applied to the chips to limit bleaching of the fluorescein. The signals were acquired by a custom built TIRF-based CCD detector.

TIRF-based image detector

The TIRF-base CCD detector consists of a: (1) set of lasers used to excite the spectrally separable dye set, such as fluorescein, Cy3, Texas Red and Cy5; (2) a mechanism for shuttering, expanding, and launching the lasers sequentially into the glass slides used as microarray substrates; (3) a filter wheel used for sequentially selecting the emission band-pass for each dye; and (4) an imaging lens and CCD imager for recording the spatial fluorescence intensities.

Light from the excitation lasers is directed along common paths. The ribbon of light strikes a continuous linearly recirculating mirror (dither mirror), which deflects the light upward toward the vertex of a prism. The cylinder lens directly in front of the prism has two functions: (1) focusing the ribbon in the narrow dimension at the prism vertex, and (2) producing a continuously changing launch angle together with the dither mirror. The range of angles produced insurers that the excitation is uniform over the slide. Fluorescence is collected by a 60-mm f/2.8 MicroNikkor objective (Nikon, Japan) at a 3:1 imaging ratio. The Quantix CCD camera (Photometrics, Tucson, AZ) is cooled to −25°C and contains a Kodak KAF-1400 CCD chip with 1,037 × 1.315 pixels that are 6.8 μm × 6.8 μm square giving maximum resolution of 20 μm over the slide in the current imaging ratio. For arrays with fewer, larger spots, the full spatial resolution of the camera is not required and the CCD pixels are "binned" in a 2 × 2 fashion permiting 4× faster imaging times. The camera, custom-machined shutters, and FW1 filter wheel (Integrated Scientific Imaging Systems, Inc., Santa Barbara, CA) are controlled by a customized version of Image Pro Plus™ software (Media Cybernetics, Inc., Silver Spring, MD), which automates the acquisition of the sequence of images required for each assay. More information about the TIRF based fluorescence detector can be found at (http://www.asper.ee).

Analysis of data

Images were processed with the Image Pro Plus software. Patterns of all four incorporated nucleotides were recorded under different color codes (A, yellow; T, cyan; C, red; and G, green). Four-color images were generated using a macro.

RESULTS AND DISCUSSION

Design of the assay

An integral part of the assay is the DNA chip. We have used standard microscope cover glasses (24 × 60 mm) activated by

TABLE 2. β-THALASSEMIA APEX SLIDE KEY

	M	*−87*	*Codon 5*	*Codon 6*	*IVS-I-1*	*IVS-I-5*	*IVS-I-6*	*IVS-I-110*	*Codon 39*	*IVS-II-1*	*IVS-II-745*	*M*
Sense	N	C to G	C to T	A to G	G to A	G to A	T to C	G to A	C to T	G to A	C to G	N
Antisense	A	G to C	A to G	T to C	C to T	C to T	A to G	C to T	G to A	C to T	G to C	C
Sense	G	C to G	C to T	A to G	G to A	G to A	T to C	G to A	C to T	G to A	C to G	T
Antisense	N	G to C	A to G	T to C	C to T	C to T	A to G	C to T	G to A	C to T	G to C	N

The left and right columns consist of self-extending marker primers (M) and the middle 10 columns are duplicates of the sense adjacent to antisense primers for the mutation sites listed in Table 1. The letters are showing analysis results expected from a wild-type DNA and mutations, respectively.

epoxy-silanization (Southern *et al.*, 1992; Lamture *et al.*, 1994) for oligonucleotide attachment. The primers were coupled to the activated slides by their 5′ amino linker under alkaline conditions. The average coupling efficiency was 5%, as estimated by immobilization of radioactively labeled oligonucleotide (data not shown).

Template preparation

We have developed a single-tube template preparation protocol, consisting of PCR amplification followed by a DNA fragmentation reaction. A 1.4-kb fragment of the patient DNA was first amplified by PCR. A fraction of the dTTPs was substituted

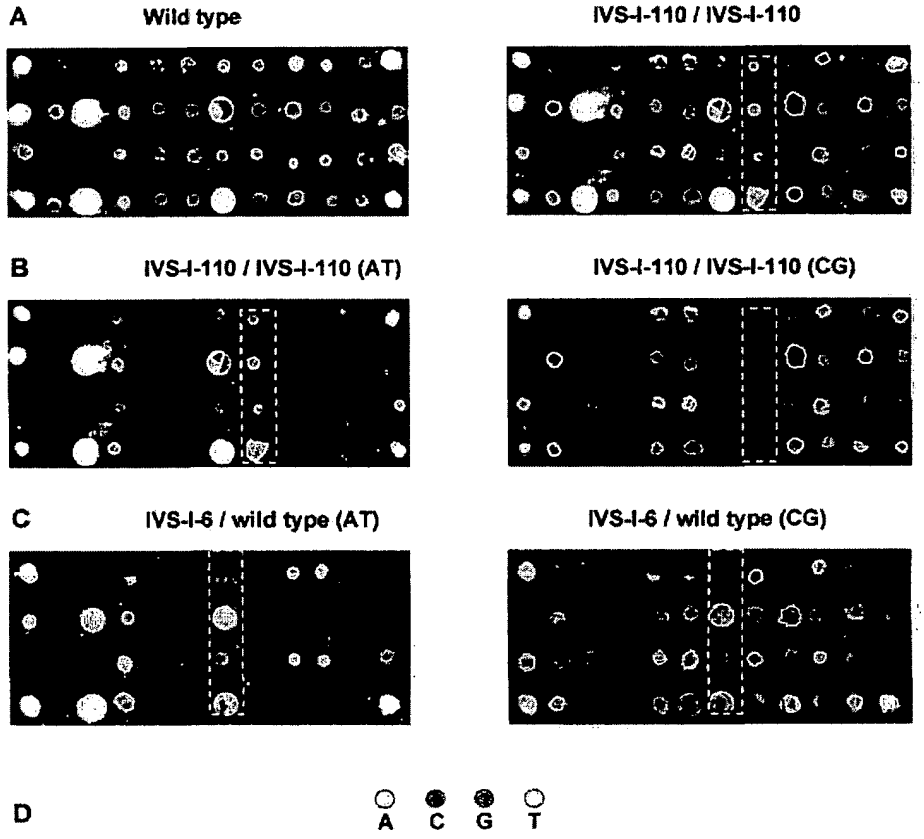

FIG. 1. Four-channel pseudocolor images of β-thalassemia APEX analysis. Table 2 indicates the positions of oligonucleotides on the array. **A.** Wild-type and homozygous IVS-I-110 (G → A; Tables 1 and 2) mutation. Dashed boxes indicate the IVS-I-110 sense and antisense primer locations. **B.** DNA sample carrying homozygous IVS-I-110 (G → A; Tables 1 and 2) mutation. Composed pseudocolor images representing signals from complementary nucleotides are shown for clarity. **C.** DNA sample carrying heterozygous IVS-I-6 (T → C; Tables 1 and 2) mutation. Wild-type allele is apparent in left panel while mutant allele is visible on the right panel. **D.** Color code of the images.

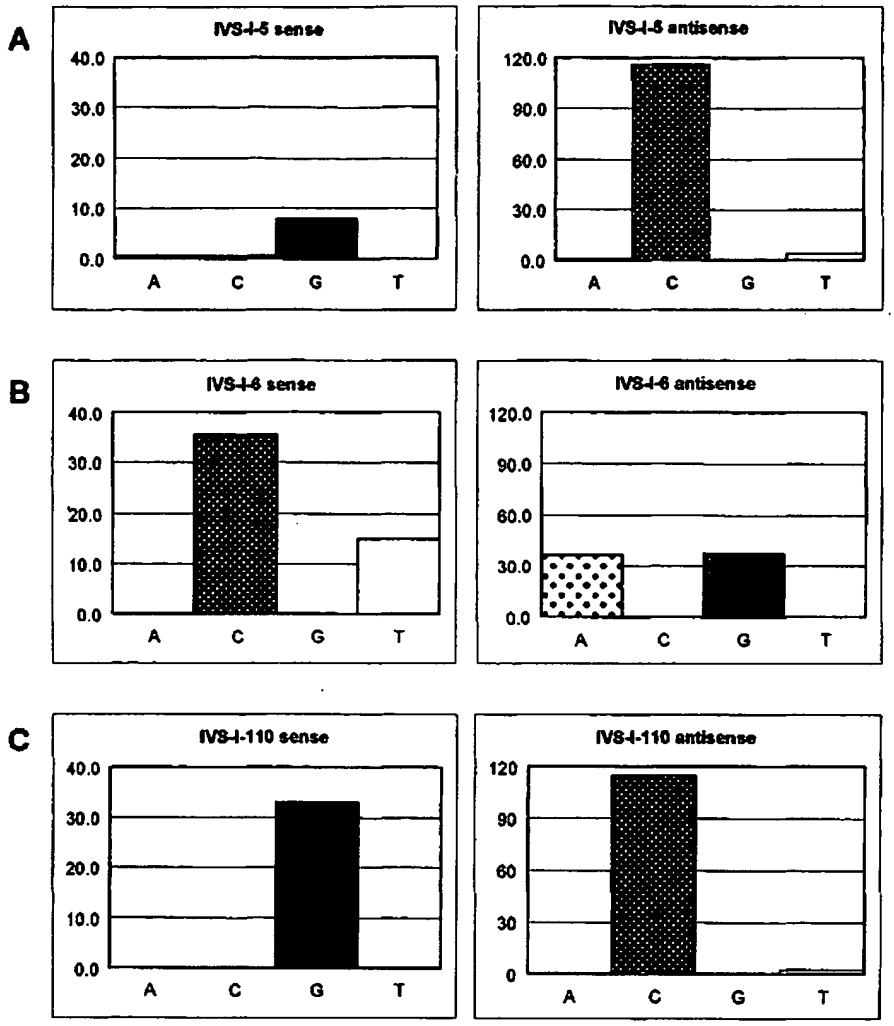

FIG. 2. Fluorescent intensities of three sequential mutation analysis sites on the β-thalassemia oligonucleotide array (Tables 1 and 2). A. IVS-I-5 site. Bars represent average fluorescent intensities of three independent experiments with a wild type target DNA. B. IVS-I-6 site. Average fluorescent intensities from two different experiments with a heterozygous target DNA. C. IVS-I-110 site showing average fluorescent intensities from the same three independent experiments as (A).

by dUTPs in the PCR mix allowing for later fragmentation with uracil *N*-glycosylase (UNG) and heat treatment. *In vivo*, UNG acts as one of the most efficient DNA repair enzymes, hydrolyzing specifically the *N*-glycosylic bond connecting uracil to the deoxyribose sugar and generating abasic sites in DNA. *In vitro*, this reaction can be used for asymmetric fragmentation of the template DNA (Cronin *et al.*, 1996). Replacement of 20% of dTTPs was optimal for β-thalassemia APEX. However, other concentrations of dUTP might be needed for templates with different lengths and thymidine content. Fragmentation offers several advantages for APEX by reducing both the effects of secondary structures, reducing the melting temperature of target duplexes, and permitting the analysis of both strands simultaneously. Fragmentation also promotes greater mobility of the template and increases its effective concentration. In addition to the UNG treatment, several other possibilities exist for DNA template fragmentation, such as DNase I treatment, restriction enzyme digestion, and mechanical shearing; however, none of these offers the combination of reproducibility, fragment size, staggered single-sided nicks, and assay flexibility.

After amplification, the PCR products were concentrated by

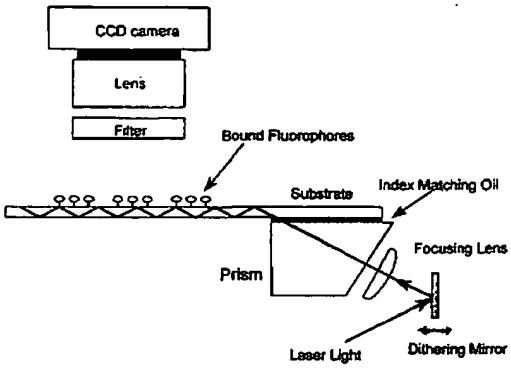

FIG. 3. Scheme of TIRF excitation/CCD imaging system. The excitation beam is trapped in the oligonucleotide array slide by total internal reflection at the slide/air interface. The launch angle is nominally 20° below horizontal. As the excitation beam travels down the oligonucleotide array slide, it expands to fill the slide uniformly. The evanescent field of the trapped beam is used to excite the dye molecules incorporated to the oligonucleotide primers near the surface. Variations in the excitation field are smoothed by the action of the dither mirror.

ethanol precipitation. dNTPs carried over from the PCR reaction are a source of nonspecific extension noise in the APEX reaction and must be removed. The remaining dNTPs were degraded enzymatically using shrimp alkaline phosphatase (sAP). Simultaneously, the amplicons were treated with UNG.

In the case of β-globin gene as the model system, amplification with a single pair of primers was sufficient to evaluate the common mutations. However, to apply the APEX approach to larger genes, amplification with multiple primers may be necessary.

APEX reaction

The majority of the β-thalassemias are caused by either a single nucleotide substitution, or an oligonucleotide addition or deletion that affects the coding region, or critical areas, for the function of the β-globin gene (Cao *et al.*, 1997). Primer extension reaction conditions were optimized with a wild-type DNA to achieve signals from all screened positions in the β-globin gene.

The specificity of the assays was monitored by immobilized self-elongating marker primers (Tables 1 and 2; Fig. 1, A, B and C) that are designed to form self-complementary homoduplexes at the 3′ end, permitting template-independent signals for particular dye terminators. All primers on the array gave signals as expected from the β-globin gene wild-type sequence (Table 2; Fig. 1A). From the analyzed patient DNA, sample 1 was homozygous for the IVS-I-110 mutation (Fig. 1, A and B) and 8 were heterozygous for the −87, Codon 5, Codon 6, IVS-I-1, IVS-I-6 (Fig. 1C), IVS-II-1, Codon 39, and IVS-II-745 mutations, respectively.

The signal-to-noise ratio of primer extension is 40:1, measured as the average fluorescence value for all oligonucleotides on the array from three different experiments (Fig. 2). This favors identification of heterozygous mutations. Covalent bonds between the oligonucleotide and dye terminator allow the slides to be stringently washed, minimizing the nonspecific signals and reducing the background.

Our goal was to make the APEX assay as simple and robust as possible; therefore, we abandoned the two-step assay described earlier (Head *et al.*, 1997; Pastinen *et al.*, 1997). To minimize manual operations with the DNA chips, we have used single-step APEX reactions. Our experiments have shown that both hybridization and template-dependent extension of arrayed primers can be achieved in the same reaction step without removal of unbound template. The same goal applies to the reaction mix; it contains only absolutely necessary components—template DNA, fluorescently labeled dideoxy nucleotides, and high-specificity DNA polymerase in its commercial buffer. Furthermore, the reaction conditions are relatively insensitive to variations in the amount of dye terminator and polymerase.

Some target-dependent primers show self-extension signals, *e.g.*, Codon 39 and IVS-II-1 sense-strand primers gave signals from the wild-type sequence (incorporation of C and G), as well as from self-extension (incorporation of A), respectively, due to formation of primer bridge structures (homodimers) similar to the marker primers used, or to hairpin structures. The primers on the array are designed according to the wild-type sequence of the analyzed gene. Although the 3′ end of the primers cannot be varied, the internal part of the primer may be changed by incorporating a mismatch to reduce primer self-complementarity without seriously affecting the target-specific priming ability. We are screening each mutation from both DNA strands. APEX analysis of the opposite strand does not use the complement of the problematic primer, and thus has a reasonable probability of avoiding the self-priming problem. In addition, there are no sites in the current assay containing such problematic primers for both strands.

TIRF detection system and CCD imaging

We have developed a detection system based on TIRF connected to a CCD image reader (Axelrod *et al.*, 1984; Stimpson *et al.*, 1995). Figure 3 demonstrates the basic idea of the excitation scheme. A laser is deflected by a mirror and focused via a lens through a launch prism and index matching oil onto a glass slide. The intensity of the light field is nearly uniform along the length of the slide. Above the surface of the slide the intensity decreases exponentially, extending for approximately one-quarter of the wavelength of light as an evanescent field, which excites the bound fluorophores residing near the surface. The emitted fluorescence is filtered to reject the background scatter noise, and adjacent dye signals are then collected by a CCD camera. Using one, two, or four dyes, four images are obtained, one for each of the four dye-labeled ddNTPs. The intensities of the imaged spots for each array element are compared and the largest signal will identify the nucleotide in the target sequence. When two signals are present at a location, a heterozygous status is indicated.

The number of detectable fluorescent labels can be adjusted according to assay requirements by the addition (or removal) of corresponding lasers and filters. Two main criteria exist for choice of dye-terminator conjugates. First, they must be spectrally separable from each other and, second, they have to be

incorporated by DNA polymerase. In this report, the "four labels-one reaction" scheme with ddNTPs conjugated to fluorescein, Cy3, Texas Red, and Cy5 was used for β-thalassemia APEX reactions.

The time required for the complete APEX analysis is less than 4 hours, including PCR and sample preparation. However, much of the target preparation and APEX reaction can be performed in parallel, and the detection presents the major limiting step in high-throughput analysis. The TIRF-CCD detector is capable of reading one four-color slide per minute, presenting an ultimate throughput of 60 slides per hour in the present design. For a small number of sites, such as the 10 sites presented here, the assay can be analyzed visually.

We propose that APEX offers a good platform for high-throughput genetic testing. The approach can be applied to any DNA target for analysis. The 40:1 signal-to-noise ratio enables identification of heterozygous mutations with comfortable confidence levels.

ACKNOWLEDGMENTS

The authors wish to thank Heidi Saulep and Viljo Soo for technical assistance. This work was supported by research grants from European Community IC15-CT98-0309, Estonian Science Foundation 2492, and Core Grant 0180518s98 from the Estonian Ministry of Education. J.S. and J.T. were partly supported by a NIST ATP award 70NANB5H1112.

REFERENCES

AXELROD, D., BURGHARDT, T.P., and THOMPSON, N.L. (1984). Total internal reflection fluorescence. Annu. Rev. Biophys. Bioeng. **13**, 247–268.

CAO, A., SABA, L., GALANELLO, R., and ROSATELLI, M. (1997). Molecular diagnosis and carrier screening for β-thalassemia. J. Am. Med. Assn. **278**, 1273–1277.

CHEE, M., YANG, R., HUBBELL, E., BERNO, A., HUANG, X.C., STERN, D., WINKLER, J., LOCKHART, D.J., MORRIS, M.S., and FODOR, S.P.A. (1996). Accessing genetic information with high-density DNA arrays. Science **274**, 610–614.

COTTON, R.G.H., EDKINS, E. and FORREST, S., eds. 1998. *Mutation Detection. A Practical Approach.* Oxford University Press, Inc., New York.

CRONIN, M.T., FUCINI, R.V., KIM, S.M., MASINO, R.S., WESPI, R.M., and MIYADA, C.G. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum. Mutat. **7**, 244–255.

HACIA, J.G., BRODY, L.C., CHEE, M.S., FODOR, S.P.A., and COLLINS, F.S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nature Genet. **14**, 441–447.

HEAD, S.R., ROGERS, Y.-H., PARIKH, K., LAN, G., ANDERSON, S., GOELET, P., and BOYCE-JACINO, M.T. (1997). Nested genetic bit analysis (N-GBA) for mutation detection in the p53 tumor supressor gene. Nucleic Acids Res. **25**, 5065–5071.

HUISMAN, T.H.J., CARVER, M.F.H., and BAYSAL, E. (1998). *A Syllabus of Thalassemia Mutations.* (ed. by Sickle Cell Anemia Foundation, Augusta, GA).

LAMTURE, J.B., BEATTIE, K.L., BURKE, B.E., EGGERS, M.D., EHRLICH, D.J., FOWLER, R., HOLLIS, M.A., KOSICKI, B.B., REICH, R.K., SMITH, S.R., VARMA, R.S., and HOGAN, M.E. (1994). Direct detection of nucleic acid hybridization on the surface of a charge coupled device. Nucleic Acids Res. **22**, 2121–2125.

LANGEGREN, U., NILSSON, M., and KWOK, P.-Y. (1998). Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res. **8**, 769–776.

PASTINEN, T., KURG, A., METSPALU, A., PELTONEN, L., and SYVÄNEN, A.-C. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res. **7**, 606–614.

SHUMAKER, J.M., METSPALU, A., and CASKEY, T. (1996). Mutation detection by solid phase primer extension. Hum. Mutat. **7**, 346–354.

SOUTHERN, E.M., MASKOS, U., and ELDER, J.K. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics **13**, 1008–1017.

STIMPSON, D.I., HOIJER, J.V., HSIEH, W.T., JOU, C., GORDON, J., THERIAULT, T., GAMBLE, R., and BALDESCHWIELER, J.D. (1995). Real-time detection of DNA hybridization and melting on oligonucleotide arrays by using optical wave guides. Proc. Natl. Acad. Sci. USA **92**, 6379–6383.

Address reprint requests to:
Prof. Andres Metspalu
Inst. of Molecular and Cell Biology
Chair of Biotechnology
University of Tartu
23 Riia St. Tartu 51010
Estonia

E-mail: andres@ebc.ee

Received for publication September 24, 1999; accepted December 28, 1999.